

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

METHOD AND APPARATUS FOR
DISCOVERING EVOLUTIONARY CHANGES WITHIN A SYSTEM

CROSS-REFERENCE TO RELATED APPLICATION

5 This application claims priority from PCT International Application No. PCT/US02/22977, filed July 18, 2002 and entitled "METHOD AND APPARATUS FOR DISCOVERING EVOLUTIONARY CHANGES WITHIN A SYSTEM", and claims the benefit of the following co-pending provisional 10 applications:

(a) Serial No. 60/374,064, filed April 19, 2002 and entitled "PROCESSING MIXED NUMERIC AND/OR NON-NUMERIC DATA";

15 (b) Serial No. 60/374,020, filed April 19, 2002 and entitled "AUTOMATIC NEURAL-NET MODEL GENERATION AND MAINTENANCE";

(c) Serial No. 60/374,024, filed April 19, 2002 and entitled "VIEWING MULTI-DIMENSIONAL DATA THROUGH HIERARCHICAL VISUALIZATION";

20 (d) Serial No. 60/374,041, filed April 19, 2002 and entitled "METHOD AND APPARATUS FOR DISCOVERING EVOLUTIONARY CHANGES WITHIN A SYSTEM";

(e) Serial No. 60/373,977, filed April 19, 2002 and entitled "AUTOMATIC MODEL MAINTENANCE THROUGH LOCAL NETS";
25 and

(f) Serial No. 60/373,780, filed April 19, 2002 and entitled "USING NEURAL NETWORKS FOR DATA MINING".

TECHNICAL FIELD

30 This application relates to numerical modeling. In particular, the application relates to adaptive system modeling and model evaluation.

DESCRIPTION OF RELATED ART

35 Numerical modeling typically involves generating a

computer program or software module which can approximate or model the behavior of a dynamic system. A system may be characterized through a system model as a group of input parameters and corresponding group of output values which 5 may be predicted by applying the input parameters. A purpose of the system model is to enable predictions or forecasts of the underlying dynamic system.

Much of numerical modeling may fall under the umbrella of machine learning. Machine learning involves 10 characterizing the behavior of a physical or procedural system, usually by utilizing a software simulation. The purpose is to enable prediction or forecasting of the system's behavior. For example, a profit may be made or a loss avoided, through a good forecast of a stock price. 15 Similarly, with other systems, some sort of risk may be avoided or benefit gained, through some foreknowledge. This is not fortune-telling, but rather a forecast based on a system's past behavior and possibly based on its past reactions to outside influences. There is some error 20 inherent in these predictions, and the goal is to minimize this error.

Machine learning is also sometimes called "blackbox modeling" because a system is usually reduced to a set of inputs and outputs, and little or no consideration is 25 afforded to the inner workings of the system.

Machine learning is different from other types of numerical modeling, such as first principle modeling, which relies on determining the inner mechanisms of a system. For complex and interesting systems first principle 30 modeling is very difficult, probably expensive, and perhaps impossible.

In the exemplary system of FIG. 4, the output is the predicted price at which a stock will close. The inputs to

the system are market factors, such as prices and activity of other stocks, and outside factors which influence the market, such as price of oil and retail sales. Even with little experience in the stock market, one may readily see 5 that the possible number of model inputs is potentially enormous. Unless the focus is on the most significant indicators of a stock's price, even a supercomputer might not accomplish the task of modeling this system.

A typical numerical modeling process includes a 10 selection of system features to treat as model inputs. Except for the simplest systems, the set of possible input features are typically reduced to a manageable set of inputs. At the least, feature reduction strives to remove features which have the least effect on the system output 15 and to remove features which appear to be static. Usually, an attempt is made to select out features with less effect on the outputs, in order to reduce the need for computational resources, as well as to reduce processing times, because manageability of the model typically depends 20 upon analysis and processing resources.

A numerical modeling process typically includes (a) a training or learning phase in which a model is generated through training from a body of known data, holding out some percentage of the data for (b) a model validation or 25 test phase in which accuracy of the trained model is tested.

A pairing of system inputs and outputs is often called a "pattern". In order to build a numerical model, a set of patterns, called a training set, is generated. In order to 30 produce an effective model, the training set should contain patterns representative of most (if not all) of a system's behavioral modes. The group of outputs associated with a particular group of input parameters in a particular

pattern may be actual outputs collected from the system when the group of inputs were applied to the system. Alternatively, the outputs in the pattern may be a group of desired outputs, when the group of inputs is applied to the system.

In the learning phase, a learning methodology, which typically is embodied in a computer program or software module, is applied with patterns in the training set, and the numerical model thereby learns the system behavior which is then stored as model parameters. Numerous learning methodologies are conventionally known. The methodologies of the present invention are not limited to application of any particular ones of the conventional learning methodologies.

The trained system model is typically tested or validated in a test phase against another set of patterns, called a "test set". The test set should not contain patterns which were present in the training set. The test phase involves, for each pattern in the test set, comparing the system output predicted by the model to the expected outputs in the test pattern. The accuracy or effectiveness of the model is gauged by how well the model predictions agree with the pattern outputs.

If the accuracy of the trained model is deemed sufficient, the model is deployed to the field and used to predict system behavior. After the model is deployed, the accuracy of the model may be periodically re-evaluated, in order to detect if the model is losing its effectiveness. If the model performance has degraded, the model may be re-trained with additional data. However, if data has not been collected continuously while the model is in operation, a large effort might be necessary to gather new data and/or to search for inputs which have now become

significant. This cycle has the additional disadvantage that a model's degradation beyond usefulness may be discovered late in the cycle, and, once discovered, the analysis and deployment steps must be repeated.

5 In some instances, new data corresponding to system inputs and system outputs may be collected and accumulated on-line while the model is in operation, and may be available for use in re-training the model. However, if re-training fails to increase accuracy of the model to
10 acceptable levels, then the feature selection performed in the original analysis must be repeated to determine if the system has evolved such that input features previously thought to be discardable have now become significant and the whole process of training and deployment also must be
15 repeated.

SUMMARY

An adaptive system modeling method is provided. In
20 one embodiment, the method includes selecting from a plurality of candidate features of a system a set of input features and a superset of the input features and other features by using a baseline significance signature, generating a system model by using data corresponding to
25 the selected input features set, maintaining online data corresponding to the superset of the input features and other features collected from the system, determining a new significance signature of the system by using the online superset data to perform a discriminant analysis of the candidate features, and detecting an evolutionary change in
30 the system by comparing the new significance signature and the baseline significance signature. The method may further comprise selecting new input features by using the

new significance signature.

The method, according to another embodiment, includes determining a baseline significance signature of current behavior of a system by performing a discriminant analysis, 5 selecting from a plurality of candidate features a set of input features and a superset of the input features and other features by using the baseline significance signature, generating a system model by using data corresponding to the selected input features set, and 10 maintaining online data corresponding to the superset of the input features and other features collected from the system. The method may further comprise evaluating an accuracy of predictions by the system model based on additional input features data, determining a new 15 significance signature of the system by performing another discriminant analysis of the candidate features, if the accuracy of the system model predictions is below a predetermined accuracy level, and selecting new input features by using the new significance signature. The 20 additional input features data may be obtained from the online collection of data.

According to another embodiment, the adaptive system modeling method includes determining a baseline significance signature of current behavior of a system by 25 using a decision tree methodology to perform a discriminant analysis, selecting from a plurality of candidate features a set of input features by using the baseline significance signature, and generating a system model by using data corresponding to the selected input features set.

30

BRIEF DESCRIPTION OF THE DRAWINGS

The features of the present application can be more readily understood from the following detailed description

with reference to the accompanying drawings wherein:

FIG. 1 shows a flow chart of an adaptive system modeling method, in accordance with one embodiment of the present disclosure, which may be applied to detect an 5 evolutionary change in a system;

FIG. 2 shows a flow chart of an adaptive system modeling method, in accordance with another embodiment of the present disclosure;

FIG. 3 shows a flow chart of an adaptive system 10 modeling method, according to another embodiment of the present disclosure;

FIG. 4 shows a schematic representation of a stock price forecasting system;

FIG. 5 shows a schematic representation of a blackbox 15 model corresponding to the stock price forecasting system of FIG. 4;

FIG. 6 shows a flow chart of an adaptive system modeling method, according to another embodiment of the present disclosure;

20 FIG. 7 shows a schematic representation of an exemplary decision tree for a coating formulation model;

FIG. 8 shows a user interface reporting training and validation results for a model trained with all available features, in connection with the coating formulation 25 example of FIG. 7;

FIG. 9 shows a graphical representation of a relationship between selected system features and an entropy measure, in connection with the coating formulation example of FIGS. 7 and 8;

30 FIG. 10 shows a user interface reporting training and validation results for a model trained with the most significant features, in connection with the coating formulation example of FIGS. 7 and 9;

FIG. 11 shows a user interface reporting training and validation results for a model trained with all available features, in connection with an exemplary wine purchase model;

5 FIG. 12 shows a schematic representation of upper levels of an exemplary decision tree for the wine purchase example of FIG. 11;

10 FIG. 13 shows a graphical representation of relative significance of features, in connection with the wine purchase example of FIGS. 11 and 12;

FIG. 14 shows a schematic representation of top levels of an exemplary decision tree for modeling a help desk system;

15 FIG. 15 shows a graphical representation of binning for a selected feature, in connection with the coating formulation example of FIGS. 7-10;

FIG. 16 shows a graphical representation of binning for a selected feature, in connection with the wine purchase example of FIGS. 11-13;

20 FIG. 17 shows a graphical representation of purchase tendencies and a selected feature, in connection with the wine purchase example of FIGS. 11-13;

FIG. 18 shows a schematic representation of an exemplary help desk system;

25 FIG. 19 shows a tabular representation of problem codes ranked by significance, in connection with the help desk system example of FIG. 14;

FIGS. 20-23 show exemplary user interfaces for configuring the help desk model of FIGS. 14 and 19;

30 FIG. 24 show a user interface for entering database connection information in connection with the modeling system of FIGS. 14 and 19;

FIGS. 25-29 show exemplary user interfaces showing

reports in connection with the help desk modeling example of FIGS. 14 and 19;

FIG. 30 shows a high-level block diagram of a help desk modeling system;

5 FIG. 31 shows a flow chart for an installation wizard of the help desk modeling system of FIG. 30;

FIG. 32 shows a flow chart for a manager service of the help desk modeling system of FIG. 30;

10 FIG. 33 shows a flow chart for a Consult module of the help desk modeling system of FIG. 30;

FIG. 34 shows a flow chart for a SavePattern module of the help desk modeling system of FIG. 30;

FIG. 35 shows a flow chart for an Evaluate module of the help desk modeling system of FIG. 30;

15 FIG. 36 shows a flow chart for a DecisionTree module of the help desk modeling system of FIG. 30;

FIG. 37 shows a flow chart for an UpdateModel module of the help desk modeling system of FIG. 30; and

20 FIG. 38 shows a data flow diagram for the help desk modeling system of FIG. 30.

DETAILED DESCRIPTION

The present disclosure provides tools (in the form of methodologies and systems) for adaptive system modeling and 25 model evaluation. For example, the methodologies of this disclosure may be incorporated in software-implemented systems to model dynamic systems. The methodologies may be embodied in one or more computer programs or software modules stored on a conventional program storage device or 30 computer readable medium, and/or transmitted via a computer network or other transmission medium.

The system modeling and model evaluation methodologies of the present disclosure reduce or eliminate a need to

begin a modeling cycle anew, when the system has evolved such that the deployed model has reached the end of its usefulness.

An adaptive system modeling method which may be applied to detect an evolutionary change in a system, in accordance with an exemplary embodiment shown in FIG. 1, includes selecting from a plurality of candidate features of the system a set of input features and a superset of the input features and other features by using a baseline significance signature (step S11). A system model is generated by using data corresponding to the selected input features set (step S12). Data corresponding to the superset of the input features and other features collected from the system is maintained online (step S13). A new significance signature of the system is determined by using the online superset data to perform a discriminant analysis of the candidate features (step S14). An evolutionary change in the system is detected by comparing the new significance signature and the baseline significance signature (step S15). A new set of input features may be selected by using the new significance signature.

An adaptive system modeling method according to another exemplary embodiment is described with reference to FIG. 2. A baseline significance signature of current behavior of a system is determined by performing a discriminant analysis (step S21). A set of input features and a superset of the input features and other features are selected from a plurality of candidate features by using the baseline significance signature (step S22). A system model is generated by using data corresponding to the selected input features set (step S23). Data corresponding to the superset of the input features and other features collected from the system is maintained online (step S24).

The method optionally may further include evaluating an accuracy of predictions by the system model based on additional input features data (which may be obtained from the online collection of data) [step S25], determining a new significance signature of the system by performing another discriminant analysis of the candidate features (step S27), if the accuracy of the system model predictions is below a predetermined accuracy level (step S26), and selecting new input features by using the new significance signature (step S28).

According to another embodiment (FIG. 3), the adaptive system modeling method includes determining a baseline significance signature of current behavior of a system by using a decision tree methodology to perform a discriminant analysis (step S31), selecting from a plurality of candidate features a set of input features by using the baseline significance signature (step S32), and generating a system model by using data corresponding to the selected input features set (step S33).

20

A. Overview

There are a variety of numerical modeling techniques available to characterize a system. The techniques range from statistical techniques to tools such as neural networks, fuzzy logic, and decision trees. The methodologies of this disclosure may be applied to any of the numerical modeling techniques. If the problem can be framed as identifying significant system inputs from a candidate set of features, then the methodologies of this disclosure may be applied.

A system being modeled may have N available input features. Through standard feature reduction techniques, and perhaps through first principles (e.g., causal theories

for all or part of the system), one may select a core group of M features to include in model inputs. In addition, some of the remaining N-M features still may influence the model output, and having P ($>M$) input features may improve 5 the model accuracy. However, the system response to the P input features may drift or change over time, as discussed below.

Selection of inputs to use from a set of candidate features and a methodology to process the inputs before 10 presentation to the modeling technique is at least as important as choosing a modeling technique (and perhaps more so).

First cuts at reducing the data may include finding features which are static or strongly correlated with other 15 features. Such features make no contribution to the modeling effort and consume resources. Including these features may also impede the modeling methodology from finding useful patterns in the data.

After removing static and correlated features, one 20 selects the features which have the most significance to the output(s) while considering the available computing resources. Usually, one is not confronted with discarding relatively highly significant features, in order to reduce the input set to a workable number of inputs. Instead, one 25 can often identify a large number of features which, at least at the time of analysis, do not seem to have a significant effect on the output(s). The remaining significant features may be used to generate a usable model of the system. For example, the system of FIG. 4 may be 30 reduced to a black box representation, such as shown in FIG. 5.

The methodologies of this disclosure utilize discriminant analysis to characterize the relative

significance of the P input features used for modeling. This ranking of features by significance is referred to herein as a "significance signature". One type of discriminant analysis is provided through methodologies for 5 generating decision trees. A by-product of tree generation is that features in the tree are ranked according to their significance to the system output.

In addition to characterizing the model by relative significance of the P inputs currently included in the 10 model, data for the N-M features are collected and maintained on-line. The group of N-M features is referred to hereinafter as a superset of the P features activated in the model. This strategy may extend to setting M to zero, and selecting P out of N possible features.

15 Models may be evaluated by consulting with data that was not used in the training process which typically provides an objective measure of how well one might expect the model to predict future data.. For the example corresponding to FIG. 4, one may train with data from the 20 immediately preceding year, except for holding out data from the most recent month. Once the model is built, one may consult the model by using the data from the most recent month.

In rapidly evolving systems, such as the stock market, 25 one might expect the model's performance to degrade over time. The model may be periodically, or even continuously, evaluated to gauge when it is degrading.

As a model degrades, it may at least be retrained on more recent data. In some instances, the input feature set 30 does not need to be changed, since only the interactions between the inputs have changed. However, in some other instances, a sufficiently accurate model is not attainable with the current set of inputs. For example, the system

may be strongly influenced by a feature not currently represented in the input set, or perhaps a feature in the input set no longer influences the output significantly and therefore may be removed. In the example of FIG. 4, new 5 outside factors may affect the market, such as an international conflict which perturbs the market's reaction to normal market forces.

The model is periodically (or intermittently) re-evaluated for its effectiveness. In addition to checking 10 for a minimum accuracy against a test set, the N-M candidate feature set is again characterized through discriminant analysis to determine the currently most discriminating features. If there has been a radical change in the significance of possible input features, then 15 appropriate features may readily be re-selected by using the new significance signature and the model is trained against the available on-line data. As in the original analysis, some percentage of the data is withheld for validation of the new model.

20 In data mining or machine learning applications, data is frequently collected continuously from the system being modeled. Thus, new data may be used for consultation of the model, to obtain predictions or forecasts. Secondarily, this data may be kept online until the 25 predictions are validated, to measure the performance of the model.

To fully utilize the methodologies of this disclosure, a superset of the currently used input features is collected. The periodic significance signatures may be 30 used to determine that other features have become more significant to the system's behavior. As different category codes become more significant, the derived features may be reshuffled accordingly.

As mentioned above, data may be collected on-line for all of the possible inputs, and feature selection may be performed as necessary, ranging over all available features. Best results might be expected by anchoring the 5 model with some fixed core group of inputs (especially under circumstances in which first principle theories for all or part of the system are available) with variation within a limited group of candidate features. The core features may be the foundation of a good model, and the 10 variable set of features enables tuning the model for evolutionary changes or drift in the system behavior.

A methodology according to one embodiment of this disclosure is illustrated in FIG. 6. Initially, the system undergoes analysis for selection and determination of data 15 pre-processing requirements. Next, the system is modeled and put on line.

However, unlike the conventional machine learning or data mining procedure, a superset of the features currently used as inputs is maintained online. Periodic predictions 20 are made and model accuracy measured. A model may be retrained with more recent data, if accuracy has degraded.

A baseline significance signature is maintained to characterize current behavior of the system. A superset of candidate input features is maintained online. For the 25 superset, the features currently included in the model are collected along with additional features not currently used for modeling. The superset may extend up to all possible input features for the model. Discriminant analysis is periodically or intermittently performed to obtain 30 significance signatures of the superset of features to enable discovery of changes in the relative influences of the available features. If a significant evolutionary change in the system is detected, new significant features

are selected as model inputs. Since the superset is maintained online, data is immediately available for training and deployment of a new model.

Even after reselection of model inputs, the model 5 might perform poorly, which might be expected in instances in which a new feature, not currently being monitored or collected, becomes significant. Frequent occurrences of poor performance may be an indication that the system was not well understood in the initial analysis phase.

10 Some conventional modeling methodologies include periodic retraining of the model. While a few conventional techniques provide for dynamic adjustment of modeling parameters prior to retraining, such conventional adjustments typically do not involve changing the model 15 inputs. Selecting of model inputs is generally considered part of the analysis phase, and is usually not performed after the model is in operation.

The methodologies of this disclosure provides substantial advantages over simply retraining with new 20 data. Use of the significance signature as a baseline, and as an ongoing measure of system drift, enables proactively adapting the system model. The significance signature obtained through discriminant analysis may be used to determine adjustments to the set of model inputs in order 25 to adapt the model to system drift.

B. Discriminant Analysis

Discriminant analysis is a methodology for identifying features which are most significant to an outcome. A 30 significance signature which provides a measure of significance of each feature results from the discriminant analysis. A change in this signature, and especially a change in the set of significant features, may indicate

evolution or drift of the system.

Any of a number of known discriminant analysis techniques may be used in the methodologies of this disclosure to determine a system's significant input features, along with their relative significance ranking. According to one embodiment, statistical correlation may be used as a simplistic discriminant analysis technique. There are also more advanced statistical methodologies for discriminant analysis.

10

1. Discriminant Analysis and Decision Trees

A discriminant analysis methodology according to another embodiment utilizes a decision tree technique to generate the significance signature. Decision trees identify features which are the most effective in splitting a data set according to a previously performed categorization of the data. As the tree is built, the most significant features (for example, in descending order of importance) are selected through the methodology, often according to information theoretical measures or statistical significance tests.

A decision tree often is used as the sole modeling tool for forecasting. However, the decision tree in some instances may not provide complete categorization of a data pattern. In circumstances in which the tree generated does not permit exact classification, a probabilistic classification also might be performed. Regardless of whether the tree permits exact classification, it is useful for determining the significant features of a system.

Certain decision tree methodologies allow for easy integration of symbolic and numeric data. Significance testing of combined symbolic and numeric features might be more difficult with purely statistical methodologies.

Discriminant analysis may include applying a decision tree methodology to a data set which has already been classified in some manner. Decision tree methodologies recursively partition the data set to obtain subgroups that 5 are separated according to their classification. At each level of the tree a test on one or more of the features is used to partition the data. The goal usually is to perform the most effective tests to obtain discriminant detection.

The primary use of decision trees may be 10 classification of unknown data. However, as an additional benefit, the sequences of tests leading to a homogenous partitioning of a subtree may be expressed as a rule (e.g., if $f_1 = 0$ and $f_2 = 1$ and $f_3=2$, then $out_1 = x$ with some probability). Such rules may enable human reviewers and 15 users of a system model to better understand how, and how effectively, the system is being modeled.

FIG. 7 shows an exemplary decision tree for a coating formulation system. Data is categorized according to thickness of the applied coating after curing. The 20 decision tree shows the most significant influences on the coating thickness, in order of their importance.

The exemplary decision tree shown in FIG. 7 is composed of nodes and branches. Node 0 is the root node, nodes 1-3 are branch nodes, nodes 4-12 are leaf nodes. The 25 population being partitioned is already classified. For the data set corresponding to FIG. 7, each pattern was classified according to whether its thickness rating was "low" or "high". The features (e.g., viscosity) were binned as "low", "med", or "high". Associated with each 30 node of the tree are counts of the class subpopulations at that node. For instance, node 0 in the coating data tree contains 25 patterns with a high thickness value, and 59 with a low thickness value. The goal is to have the leaf

nodes be homogenous, and thus completely partition the starting population. However, the leaf nodes often have only a preponderance of a single class. Sometimes there is no clear majority, because the tree growth procedure was 5 deliberately stopped or no test exists for the data to further partition the node. In the example shown in FIG. 7 only a probabilistic classification or rule is available for this branch of the tree. Nodes 4, 5, 6, 8, and 12 in FIG. 7 are examples of non-homogenous leaf nodes. From 10 FIG. 7 the following rules may be drawn: (a) surftens and viscosity are the most significant features; (b) formulations with low surftens and medium-to-high viscosity usually form thick coatings; (c) low surftens and low viscosity usually lead to thin coatings; and (d) medium to 15 high surftens usually lead to thin coatings.

The rules may be checked against a domain expert's knowledge or intuition of how these formulation parameters affect the likely thickness of an applied coating. Some of these rules may be unfamiliar to the domain expert, or they 20 may be called into question. These may be verified, and if true represent discovered knowledge about the system.

The tree is grown by choosing a test at each level which suitably partitions the data set. Usually each possible test is tried in turn and the best one chosen 25 according to a diversity function. Individual diversity functions are discussed below. The diversity functions attempt to measure how well the feature test cleaves the data set. Most methodologies choose the same test across the entire level, in which case a selected feature test is 30 used only once. Evaluations of feature tests applied further down in the tree are less reliable because they rely on subpopulations which are progressively less representative of the entire population.

Decision tree methodologies may be roughly broken down into binary and non-binary methodologies. Binary methodologies only allow two branches per node (e.g., 0 or 1, Yes or No, etc.). Non-binary methodologies allow many branches per node, and lead to bushier trees. The non-binary methodologies are perhaps better suited for data which is not fundamentally binary in nature. However, almost any data may be binned into two bins or derived fields created to make the data appear binary. Since non-binary trees tend to be bushier, they quickly arrive at the point at which each split is less reliable as a discriminator. In addition, an effect sometimes found in non-binary trees is that some feature test can fracture the data set into many small, nearly homogenous, nodes. An example might be testing on a person's name, which might produce leaf nodes for every person in the database. Such a test is unlikely to produce a useful rule about the system.

The methodologies use a finite number of labels or ranges for the possible feature values. The methodologies may be used against continuous numeric data by means of discretizing or binning the values. As a first approximation, an arbitrary number of evenly spaced ranges is selected for each numeric feature, and the feature value is labeled by the range, or bin, it falls in. For instance, the total observed range of the feature may be broken into 3 equal intervals, and labeled low, medium, and high. Some researchers recommend further optimizing the binning operation. One methodology is to linearly cluster the values and let the data choose the best bin ranges. Alternatively, the binning operation may be combined with the feature test evaluation.

Decision tree methodologies also utilize a process

called pruning to reduce the complexity of the tree. Some of the methodologies utilize a validation set to gauge the effectiveness of pruning. Some use only the training data to identify branches to be pruned. It may be largely unnecessary to hold back data from the training set to validate pruning. A pruning methodology may be as simple as specifying a minimum node count. Pruning may also be accomplished by attempting to merge branches of a tree into a single branch through conjunctions and/or negations of feature tests. Some researchers believe complex trees should be pruned until they are compact and readily understandable to humans. Others point out that complex systems may lead to seemingly complex decision trees.

Decision tree methodologies are examples of supervised learning. These methodologies are subject to the same hazard of overtraining that plagues other supervised learning methodologies. Overtraining amounts to learning signatures of the training patterns, rather than learning underlying relationships between the training patterns. An overtrained model may adequately predict outputs for patterns in its training set, but does not generalize to predict well for patterns which were not in the training set. In order to minimize the risk of overtraining, trees may be grown using a training data set and tested using a test set, or may be pruned to remove branches that are idiosyncratic to the training data. Alternatively, some type of stopping condition may be applied.

2. Exemplary Applications of Decision Trees

The tree in FIG. 7 resulted from constraining the tree to only two levels, to keep the tree and related rules relatively simple. However, even with this constraint, it is possible to make relatively strong statements related to

the two most significant features. The following features are included in the example corresponding to FIG. 7: viscosity, surftens, persolids, nipclear, ra, rp, rl. These features are chemical properties of a coating 5 formulation and are used to predict the thickness of an applied coating.

Neugents® technology provides unsupervised and autonomous agents, which are very good at modeling and predicting behavior and at predicting probabilities of 10 transition from one system mode to another. Neugents can detect system behavior patterns which most humans cannot.

Two types of Neugents, Value Prediction and Event Prediction, might be utilized.

An Event Prediction Neugent is trained through 15 clustering of historical data of a system's behavior, to look for specific types of events that are defined in terms of values and logical conditions in the input data. For example, an Event Prediction Neugent object is configured with the names of events to be predicted and definitions of 20 the events in terms of the input data. The Neugent is then trained from historical data. In production, the Event Prediction Neugent is called to make an evaluation when provided new data, and the output is either a detection of a defined event in the data, a detection of an unusual 25 situation, a prediction of an event, a probability of an event occurring within a future time period, or an alert that the system is behaving unusually compared to the historical data from which it was trained.

A Value Prediction Neugent is an implementation of 30 neural nets for predictive modeling. A Value Prediction Neugent object is trained with a training set obtained from known input-output pairs from properly normalized independent and dependent historical data. In production,

the Value Prediction Neugent is called to predict the values of dependent features when provided the values of independent features. For example, incoming data is passed through a Value Prediction Neugent's neural network model, 5 and the model estimates the value of an unknown variable. Some examples of value predictions include predicting inventory levels and predicting whether a client might buy a particular product.

Decision tree modeling may be an adjunct methodology 10 to Neugent Value Prediction and Event Prediction models. Decision trees may assist a Neugent technician to find and include the significant features in the system model, and remove features which have little or no effect on the outputs to be predicted, to produce efficient and compact 15 system models.

In the coating formulation model, there were 84 patterns in the training set and 9 patterns in the test set. Results obtained by using all the available features are shown in FIG. 8.

20 Applying discriminant analysis to select a feature set allows a reduction in the number of features for developing a good model. The significant features identified for the training set are, in order of significance, surftens, viscosity, persolids, rl. FIG. 9 shows the relative effect 25 of each feature in characterizing the coating thickness. The relative effectiveness is measured by the drop in a measure called entropy. A Value Predict Neugent may be trained using only the four most significant features, which achieves results nearly as good as with the full set 30 of features available. FIG. 10 shows the R^2 values obtained for the reduced feature set. There were no strong statistical correlations between any of the features in the full set.

In this example, the full set of available features produced slightly better R^2 values. However, the R^2 values for the reduced feature set are very acceptable and the model using the reduced feature set might be expected to 5 generalize better on patterns not yet seen. In addition, use of discriminant analysis may produce rules which to some extent provide a view of inside the Neugent black box model. This type of insight into the relationships underlying the data helps the domain experts to validate 10 and accept a Neugent model.

A second exemplary application uses data showing purchasing patterns in a store. Each pattern of the data contains purchased quantities of selected products. The goal is to identify what combination of product purchases 15 are an indicator of a likely wine purchaser. With this knowledge, a wine retailer may target a sales promotion at customers who are most likely to purchase more wine. Data are present for 17 different products, versus the purchase of a specific brand of wine. FIG. 11 shows the results 20 obtained using all 17 products for predicting wine purchases. FIG. 12 shows the upper levels of a decision tree for the wine data. FIG. 13 shows the relative significance of the products in the wine data.

Since the output is binary (i.e. whether wine was 25 purchased or not), the R^2 is not the best measure of the effectiveness of this model. Instead a threshold is applied to the floating point output. Output values above the threshold indicate a wine purchase. Applying a threshold of 0.6 to the output of this model yields a 30 training accuracy of 94.4% correct, and a test accuracy of 88.9%. The model made 1 wrong prediction out of 9 patterns in the test set.

Using only the four most significant features another

Neugent model was trained. Upon applying the threshold to the Neugent outputs, a training accuracy of 93.3%, and test accuracy of 88.9% was obtained. The Neugent model using the full set of features and the model using only the top 5 four features each produced the wrong result for the same test pattern. Accordingly, a reduction from 17 to four features yielded essentially the same model performance. Two of these features were completely static and they were eliminated by conventional pre-processing. However, there 10 were no strong correlations between the remaining features.

The relative unimportance of many of them are not readily apparent.

A third application involves using discriminant analysis to help in feature selection for Neugents to model 15 performance of a Help Desk system. The goal was to predict whether a help desk call is likely to be solved in the time allotted according to a Service Level Agreement (SLA).

Based on general knowledge about the nature of the data, various direct and derived features were compiled 20 into patterns. Neugent modeling using the full feature set produced relatively poor results. Applying discriminant analysis produced a reduced set of significant features which produced acceptable predictions. Since the desired output of the model is binary prediction, a threshold is 25 applied to the continuous valued output to constrain it to 0 or 1 (i.e. SLA violation or not). Again, the R^2 value is not useful in evaluating this type of Neugent model. After applying a threshold, the accuracy of the model on the test set was in the range of ~90% correct. The second most 30 significant feature identified was the customer code, which might be expected to have a strong effect on SLA violation rate. However, there was an interesting aspect related to the most significant feature. This feature is the day of

the week that the call might have gone into violation. FIG. 14 shows the top levels of a decision tree for the help desk data.

The performance of the help desk for a specific call apparently varies depending on the day of the week the call is made. The overall violation rate is about 61.7%, but it varies considerably over the course of the week, as follows: Sun - 100%; Mon - 70.2%; Tue - 85%; Wed - 49%; Thu - 45.7%; Fri - 41.5%; Sat - 92.3%. It is highest in the beginning of the week and somewhat high towards the end of the week. A possible explanation is that there is more time and/or resources available to service calls in mid-week.

15 3. Additional Decision Tree Methodologies

ID3 (Iterative Dichotomiser 3) is a non-binary method which uses an information theoretic measure of information gain for a diversity function. C4.5 and C5.0 are successors of ID3 which utilize an entropy measure and provide additional strategies for pruning trees and for limiting tree growth.

Information gain is the decrease in system entropy following the application of a selected feature test. Entropy is a measure of disorder or non-homogeneity within a system. Each node at a selected level of the tree has a certain entropy and the node entropies are combined as a weighted sum to yield a system entropy for the tree level.

At each level the test which provides the largest decrease in system entropy, or information gain, is selected as the best possible test at that level.

C4.5 and C5.0 use a measure called gain ratio instead of information gain as the diversity function. Gain ratio is the information gain from a feature test divided by the

number of subsets created by that test. This measure penalizes the type of test mentioned above, which fractures the data set (e.g., person's name). If there is another test which provides a reasonable information gain with far 5 fewer subsets in the partition, it is chosen instead.

In regard to pruning, ID3 utilizes a bonsai type pruning technique of simply stopping tree growth based on a minimum node size, or specifying an acceptable minimum entropy at which to quit testing. C4.5 and C5.0 provide a 10 strategy for pruning a branch from the tree based on Bernoulli's binomial formula. According to the C4.5/C5.0 pruning strategy, an error rate is determined for each node, by designating the node as a leaf node for the class which predominates in that node and determining the number 15 of items which are mis-classified. It then assumes the observed error rate is overly optimistic and substitutes a more pessimistic error rate, using the binomial formula and a selected confidence level. If a branch yields an increase in error rate after the pessimistic substitution, 20 it is pruned. C4.5 and C5.0 use the training data for pruning as well as training. Utilizing the training data to implement pruning might be efficient. As an alternative to using the training data, one might consider using test data along with the binomial formula to implement pruning.

C4.5 and C5.0 also may use a grouping function for 25 combinations of feature tests at each level. This may detect synergistic effects whereby two tests applied simultaneously are more effective than each applied singly.

C4.5 and C5.0 employ a separate module to generate 30 rules from the decision tree. For other methodologies this might be considered another type of pruning, whereby multiple branches may be merged into a single branch through conjunctions and negations of feature values.

Chi-squared Automatic Interaction Detection (CHAID) is a methodology which differs from ID3, C4.5 and C5.0 in a number of ways. CHAID uses the Chi-squared function as a different diversity function. When applying the Chi-squared diversity function, CHAID stops tree growth before overfitting occurs. Therefore, pruning after the tree is built is unnecessary. In addition, it deals with strictly categorical variables. Accordingly, continuous variables are binned prior to beginning the tree generation procedure. ID3, C4.5 and C5.0 may use dynamic binning of continuous variables, perhaps based on information gain. Like ID3, C4.5 and C5.0, CHAID is a non-binary methodology.

CHAID compares the relative effectiveness of each possible feature test. The Chi-squared test, a statistical significance test, is used in a number of ways in the CHAID methodology. The number of branches allowed by a feature test are limited to those which produce statistically significant splits. Further, when comparing the results of each feature test, splits which produces the most significant results according to the Chi-squared test is the winner at each level. The Chi-squared test is used to halt growth of the tree when further feature tests stop producing statistically significant results. This last determination involves the use of a significance threshold, and this is the main tuning parameter available with the CHAID methodology.

Exhaustive CHAID is a variation on CHAID. When evaluating the possible branching factor for each feature test, it progressively merges the possible branches into two groups that are most self-similar according to the Chi-squared test. This basically constrains the tree to become binary.

Classification And Regression Trees (CART) is a binary

tree methodology. Like the ID3/C4.5/C5.0 family, CART may use different diversity functions to measure the effectiveness of each feature test. The following diversity functions may be used:

- 5 (a) $\min(P(c_1), P(c_2))$, where c_1 and c_2 are the two possible result classes for the test, $P(c_1)$ and $P(c_2)$ are the relative frequencies of c_1 and c_2 ;
 (b) $2*P(c_1)*P(c_2)$; and
 (c) $[P(c_1)*\log(P(c_1))] + [P(c_2)\log(P(c_2))]$, also called the entropy or information gain measure.

10 The diversity function is evaluated before and after each feature test, and the one which provides the largest decrease in diversity is declared the winner. This proceeds much as the other methodologies until the feature tests are exhausted, or until an arbitrary stopping point 15 is reached (e.g., minimum node count).

15 CART uses an adjusted error rate function $[AE(t) = E(t) + a*\text{LeafCount}(t)]$ to generate a pool of candidate subtrees. The first candidate is selected as follows. The 20 adjusted error rate is calculated for the possible subtrees containing the root node, as a parameter is gradually increased. When the adjusted error rate of a subtree becomes greater than that for the root node, then that subtree is pruned. The second candidate is chosen by 25 repeating this process starting with the first candidate subtree. The process continues until only the root node remains. A validation set of data, which was not used in the training data, is used to select among the pool of pruned candidate subtrees. The subtree with the lowest 30 overall error rate on the validation set is declared the winner. Sometimes a cost function (e.g., some weight multiplied by the probability of misclassification) is applied along with the error rate to evaluate the best

subtree. A third test data set, which is exclusive of the training and validation set, may be used to gauge the prediction or classification capabilities of the final subtree.

5

4. Binning considerations

Except for data which is completely discrete, binning is a pre-processing step which may be performed for the above described methodologies. The technique is to pick n bins, with equally spaced ranges throughout the observed range of the data. As mentioned above, it is also desirable to minimize the number of bins to avoid quickly fragmenting the data into numerous non-representative subgroups.

15 For the data shown in FIG. 15, corresponding to the coating example above, this type of linear binning might be suitable. FIG. 16, corresponding to the Wine example, shows feature data which might be considered more exponential in appearance.

20 Linear bin limits were used for the coating and Help Desk applications. For the Wine example, account was taken for the exponential appearance of most of the feature data, and purchase quantities above a certain value were binned together. This exponential appearance may be 25 characteristic of purchasing data, as purchasing large quantities of consumer goods is relatively rare.

One might expect other distributions, particularly the normalized distribution (i.e. the bell-shaped curve), to occur. FIG. 17 shows the distribution of Store Brand purchases in the Wine example. Fitting this data to a normalized curve yielded a mean of about 15 purchases of store brands, with a standard deviation of about 8. This feature may produce into three bins - bin1: quantity

purchased less than the mean minus 1 standard deviation (i.e. 7 items or less); bin2: quantity purchased within the mean +/- 1 standard deviation (i.e. 7 to 21 items); bin3: quantity purchased greater than the mean plus 1 standard deviation (i.e. 22 or more items). An automated binning tool may take account of the distributions present in the individual feature data, and bin accordingly.

Meta-data or *a priori* knowledge may also be used, when it is available. For instance if a feature is known to be 10 a date, it might be useful to represent the date as a day of the week and bin it accordingly.

5. Pruning/stopping conditions

As mentioned above, tree-pruning is crucial to 15 applying decision trees independently. The dependence of decision trees on binning and pruning considerations are some of the reasons, they are considered complementary tools to standard Neugent technology. Pruning methodologies may be utilized, even as an adjunct to 20 Neugents.

It often occurs in the tree building methodologies that a feature test which is very effective for a branch of the tree (i.e. for some sub-population of the universe) is not selected as the most effective for the entire level of 25 the tree. This has the result that locally effective tests may appear less significant than they possibly should be. Or their significance may be masked entirely. This also has the result of clouding or obscuring any rules related to that branch of the tree. This effect might be mitigated 30 by various of the pruning procedures. Another alternative to pruning is selection of multiple features at each level of the tree, in order to find optimal feature tests for each sub-population. However, it likely increases the

computational complexity of the methodology.

Other types of stopping conditions may be used as well, such as maximum tree level, or for ID3 a minimum entropy or minimum entropy drop. For on-line applications 5 in which time may be at a premium, a maximum time to process a tree level, or the total tree, may be specified.

6. Efficiency and large-scale data set considerations

Each of the methodologies use a pass through the data 10 set for each feature test which is evaluated. The binary and non-binary ones cannot be directly compared, as applied to a given data set, because they have different numbers and types of features. The effort of arbitrary binning or adaptive binning and feature test evaluation are comparable 15 for the methodologies discussed above.

Comparison of pruning operations is less straightforward. CHAID, which eschews pruning altogether might seem to be the winner here. Using limited pruning with C4.5/C5.0 appears to be less calculation intensive 20 than CART. However, if the effort to apply C4.5/C5.0's rule generation module is expended, their total pruning load are comparable. Although C4.5/C5.0 has the benefit that it uses the training data to validate pruning, CART reads a test set or holds back training data.

25 If one assumes that many of the modeled systems contain continuous feature variables, then the effort to re-frame the features to a binary format provides an edge to the non-binary methodologies.

Sufficient statistics is a table of counts of 30 occurrences and co-occurrences of the possible features. This is performed to minimize the amount of information held in RAM while the feature tests are recursively applied. This methodology has the following two side

effects. If one is dealing with continuous variables, or in the case of binary trees with ordinal valued features, in order to generate bin counts, the bin ranges already are decided upon. Any forms of dynamic binning (e.g., to 5 maximize information gain) are forgone. Also, one loses track of what individual patterns in the data set contributed to the formation of a selected branch of the tree. Once confidence has been gained in an applied methodology, this knowledge may be only of academic 10 interest however. Another factor not considered in this single-pass methodology is how to determine the ultimate range and possibly scaling of the feature values. With a suitably indexed database, minimum, maximum and possibly other statistics might be obtained with a query, without 15 the cost of a total database scan. For other data sets, this amounts to another scan, or the data is held in memory. If scaling and binning can be settled independently of the decision tree methodology however, this may be an attractive option for large data sets.

20 As a practical matter, one might be faced with a finite number of passes through a database, to handle the effects described in the previous paragraph. Two passes, with the first to gather bin statistics, might be sufficient in most applications.

25

7. Decision Tree Products

There are numerous software vendors or other sources, which supply software or function libraries implementing one or more of the above decision tree methodologies. 30 Listings of such sources may be found on the web.

Some of the factors in the selection of a decision tree tool are the following. The decision tree tool is used as an adjunct methodology and not as the principle

means of modeling a system. However, of the commercial decision tree applications typically are used as the sole or primary means of system modeling. A modeling tool typically has features to account for various distributions 5 in its binning operations.

A second factor is that real-world applications very often involve continuous valued data. The effort to recast problems to a binary-valued format may be motivation for adopting one of the non-binary tree methodologies.

10 Since real-world data mining applications may involve very large amounts of data, a modeling tool ideally enables a single-pass through a database to gather a reduced set of sufficient statistics, as discussed above. When working towards a single-pass tool, binning parameters are decided 15 before, or early in, the scan process. However, a modeling tool may include options to use both static and dynamic binning.

C. Exemplary Help Desk application

20 FIG. 18 illustrates a Help Desk system. A fictional ACME Company contracted their computer support to an outside service. To ensure a high quality of service, a SLA was put into place, committing the outside service to resolve reported problems within 24 hours.

25 The Help Desk system may be modeled, by incorporating the methodologies of this disclosure, to predict whether a call is serviced within the 24 hours required by the SLA. An Advanced Help Desk (AHD) software tool may be provided to assist managers and call center technicians in 30 anticipating problems that may adversely affect customer service levels.

Various features were identified through an AHD tool to have an influence on whether a call will be serviced in

a timely manner. Examples of features might include the group or customer reporting the problem, the priority of the problem, etc. One feature which may be significant is the problem codes associated with the call. Thousands of 5 possible codes may be defined in the problem tracking system, hundreds of which were actually represented in the data. Including all possible, or even probable, codes in the modeling effort may be inefficient. In addition, one might expect that over time the mix of problem types may 10 evolve. The solution is to model on codes which are most significant to SLA violations, at any particular time. Discriminant analysis is used to select the top P most significant codes and include them along with other significant features. A baseline significance signature of 15 problem codes is maintained.

As outlined above, model accuracy is periodically measured and the model is retrained, if necessary. Next, the significance signature is regenerated, and the codes to model on adjusted as necessary. FIG. 19 illustrates an 20 example of how the mix of problem codes may change over time. The presence of new codes at a higher ranking probably signifies a shift in the types of problems being reported to the Help Desk and a corresponding change in how well the Help Desk can resolve the current issues.

Inclusion of these codes in the modeling effort, while 25 not being the most significant features for this system overall, enabled the required prediction accuracy to be achieved with the system model. Utilizing the significance signature enabled quick detection of system evolution and model adaptation and retraining. This is an example of 30 using the significance signature to tune a model's usage of a small but volatile part of a system. Also, a superset of possible features was not maintained online in this case

because the input features involved are derived from a single field.

An AHD tool, according to an embodiment of the methodologies of the present disclosure, is described below 5 with reference to FIGS. 20-38. A block diagram of the AHD tool is shown in FIG. 30.

Neugents may be included in the AHD tool for monitoring and forecasting call levels to assist managers in forecasting system and queue loading, days and weeks 10 ahead resource planning and balancing resources. Neugents also may be provided to monitor individual open calls, and, based on the historical behavior patterns, identify calls at risk of violating the SLA and alert managers and users to the calls. The individual Neugents are adapted to 15 evaluate evolutionary changes in system behavior. Many of the behavior patterns are too complex for humans alone to identify violation trends. The AHD tool also helps managers and users to reassign or reprioritize such calls.

20 Installation

An installation wizard (FIG. 31) may be included to perform various setup operations which are only performed once or a few times. An installation wizard may install templates which allow users to choose an issue model with 25 which the Neugent works. Users may map an issue to Requests (e.g., user problems or issues being tracked in the AHD system) alone, to Change Orders (related to corresponding Requests) alone, or to a combination of Requests and Change Orders. Using the installed templates 30 users can customize the Neugent to monitor individual queues and sites within AHD.

Configuration Utility

A configuration utility may be provided to enable selection and/or management of various system parameters and definitions.

A configuration utility may provide multiple types and instances of Neugent configurations. For example, basic types of Neugents may include CallLevel Neugent and SLAViolation Neugent. Typically, multiple instances of CallLevel Neugents are possible. Multiple instances of SLAViolation Neugents may exist as well. For example, 5 wizard(s) may be provided to perform common tasks, such as deploying new Neugents instances (e.g., a Neugent dedicated to a particular queue). In addition, export mode configuration may be provided to define new styles and types of Neugents or to customize standard templates, such 10 as the following: selection of Neugent instance intervals, etc.; feature set and related queries to include in pattern generation and modeling; association of AHD Events with a Neugent instance; train window (e.g., age of data to include in new models); and data aging (e.g., age at which 15 to purge data). A standardized feature set, and related stored queries, may be included with installation of AHD. The configuration interface may provide advanced customization of queries and feature sets when desired. 20

After selecting a Neugent instance to operate on, a tab-style interface similar to that in FIGS. 20-23 may be provided to allow entry and/or modification of the collections and properties associated with the Neugent instance. FIG. 24 shows a sample interface to enter database connection information for the AHD database. The configuration module may enable utilizing stored database procedures, in place of dynamic database (e.g., SQL) 25 queries, if and when appropriate. The configuration module may provide tools to import/export stored queries from/to

script files through the Tools menu. In addition a Test query button may be provided for viewing and/or executing the dynamic SQL query.

5 Manager Service

A manager service (FIG. 32) may be provided for the following functionality on a schedule determined by the properties of an object definition for each Neugent instance defined on the system. If the Manager Service 10 does not find a Neugent model for an active Neugent instance, it switches to Historical Mode, in which available historical data is gathered, and a provisional model is trained and deployed.

Once there is an active Neugent model the manager 15 service, at a specified interval (e.g., nightly), builds a collection of consult patterns from on-line data using the specifications defined in the Managed Object store. The collection is presented to the Neugent Provider for consult against the active model. If conditions warrant (e.g., 20 predicted SLA violation) then the events specified in the object definition are raised through normal AHD channels, and reported through the AHD Historian.

Following consult, a process to build and store 25 patterns for items which have closed is launched. The conditions for item closure are specified in the object definition.

At specified intervals (e.g., nightly), an Evaluation 30 process is run to consult patterns which were withheld from the training process. Summary statistics on the results are saved and made available for viewing through the AHD Historian. During the Evaluation process, a DecisionTree process is run, according to specifications defined in the object definition, which may identify that the model should

be retrained using different codes as inputs because the mix of significant category codes has changed. If the Evaluation process determines the existing model is inadequate, either due to insufficient accuracy or due to 5 evolution of the system as detected by DecisionTree analysis, the manager service initiates retraining of the Neugent model.

Features are provided to efficiently schedule the sampling (and other) phases of each Neugent instance, to 10 avoid overlapping resource intensive phases (e.g., Training) of multiple Neugents instances. Large rollup queries may be run in off-peak times (e.g., Lunchtime, nighttime).

For each Neugent instance the manager determines the 15 appropriate methods to invoke for the current system conditions. Each of the methods below maps to the similar method for a given Neugent instance.

Upon a Neugent instance becoming active, the Manager Service switches to Historical Mode in order to gather any 20 available historical data. Historical Mode uses query conditions specified in the Neugent instance definition to gather the data in a backward-looking fashion. Following collection of historical data, a Neugent model is trained and deployed, to enable immediate prediction capability 25 when historical data is available.

Consult Method

When there is an active model for a Neugent instance, the manager service at the specified sample interval for a 30 Neugent instance, builds a pattern(s) for the Neugent by using the feature set specifications defined for the instance. A Consult method (FIG. 33) uses the FeatureSpec definitions for the Neugent instance to generate one or

more dynamic SQL queries. If preprocessing of the SQL results are specified, then such preprocessing is performed. Queries and subqueries which only generate output features for the model are not performed for 5 consult.

As an alternative to using dynamic SQL queries, stored query procedures may be used. Parts of a query which are dynamic are parameters for the stored procedure.

The consult module interprets FeatureSpec and 10 PreprocessingSpec tokens to perform tasks such as invoke an AHD function call to transform an SLA_expiration_time into an expected failure date, and then transform the date to an expiration_day_of_week and to expiration_time_of_day features. The specifications for preprocessing are 15 contained in the object definition for the Neugent instance.

The consult result is compared against event triggers for the Neugent instance. For any triggered events, the appropriate actions specified in the event definitions for 20 the Neugent instance are taken. The actions taken in many instances are to raise AHD events, which generate alerts and notifications through normal AHD channels. Neugent calls the RaiseEvent utility when conditions warrant, and AHD determines when and how to alert users of a Neugent 25 prediction.

Due to the nature of the dynamic SQL queries being built, the Consult module has a debug or expert mode enabled by, for example, a Windows Registry entry. The mode enables logging of some or all of the queries being 30 generated and submitted to the database server.

SavePattern Method

Patterns are stored online through a Consult method

(FIG. 34) to avoid having to do a costly or perhaps impossible reconstruction process offline. For example, if a feature value which is germane to an SLA Violation changes some time after the issue went into violation, it 5 is difficult to recover that value without traversing database transaction logs. In addition, for rollup type features, it might be difficult to reconstruct a historical pattern.

The SavePattern module performs similar operations to 10 the consult activity, except that the ClosureConditions in the object definition for the Neugent instance is evaluated and output features are gathered as well. The final results are combined to form a pattern which is stored in the local namespace framework. These patterns are later 15 used for DecisionTree analysis and/or re-training. The pattern is likely a superpattern, in which the current model is not using all the available features. In that case additional features are being stored in the event it becomes desirable to use a different set of features in the 20 model.

The replaceable tokens in the dynamic SQL queries (or stored procedures), in order to build patterns, include possible references to values in the local data store which allows, for example, an SLAViolation type Neugent to use 25 the hourly sample of a CallLevel type Neugent for something like "calls opened in the last hour" as part of its pattern. A costly rollup query need not be repeated as frequently as the SLAViolation Neugent samples (e.g., 10 minutes) and instead the hourly average from a companion 30 Neugent may be used.

The module creates a table in the local namespace framework, if one does not exist already.

An end of day function for the pattern build module is

to purge pattern data beyond an aging limit specified in the system configuration. A time of day property in the object definition for the Neugent instance allows for synchronization with normal backup procedures.

5 Due to the nature of the dynamic SQL queries being built, the module has a debug or expert mode enabled by something such as a Windows Registry entry. The mode enables logging of some or all of the queries being generated and submitted to the database server.

10

Evaluate Method

An Evaluate method (FIG. 35) is called at intervals (e.g., Daily) specified in the Neugent instance definition.

15 If a model is active for the Neugent instance and data is available, summary performance statistics are generated and stored for the instance. The data is available for use by the UpdateModel method, or for viewing and reporting with the AHD Historian. Performance evaluation does not rely on R² values as a figure of merit. Instead the figure of merit 20 is tied to the Event conditions for the Neugent, i.e. whether an Event is raised when called.

DecisionTree Method

A DecisionTree module (FIG. 36) utilizes the 25 DecisionTree parameters for the Neugent instance to perform discriminant analysis on available data. The primary candidates for re-evaluation in the AHD feature set are AHD category codes. The codes improve the effectiveness of a Neugent model. However, since the codes have no natural 30 numerical ordering they are included in a Neugent model as binary features. Including non-significant features may adversely affect the Neugent model performance. Further, there may be too many features to include all in the model.

Therefore, only the currently most significant ones may be included.

The category codes are extracted through significant preprocessing. By default, discriminant analysis is only 5 enabled against the category codes. However, there are properties in the Neugent object definition structure to include other features in a DecisionTree analysis.

UpdateModel Method

10 Under conditions specified for the UpdateModel instance (e.g., time of day) [FIG. 37], the manager service checks for availability of data for DecisionTree analysis and/or Neugent training. If no or insufficient data is available the model does nothing further.

15 If called by the Object definition for the Neugent instance, DecisionTree discriminant analysis is performed for available data. If the significant features have changed from what is in the current feature set, then the current Neugent instance object definition is backed up. 20 Next, a flag is set to indicate a train set is to be preprocessed according to the new significant features, and re-training is initiated as discussed below.

If the Neugent instance calls for unconditional re-training at specified intervals, the method initiates 25 training of a new model, with random validation, provided there is available data.

If the model meets minimum standards for accuracy, the manager attempts to install it as the current model. If there is no current model for the instance, the manager 30 installs the model unconditionally. If there is an existing model, the manager compares the effectiveness of the new one to the currently active one. If the new one is more effective, it is marked as the active one. Otherwise,

the old one stays active.

Historian Module

A Historian interface for user reporting (e.g., summary reporting of call-level prediction Neugents) may be provided. The Historian module provides an interface for users to view or print reports of Call Level Neugent predictions and/or prediction effectiveness. For example, users may select from the available Neugents instances and types, and then view available prediction and performance reports for the selected type. While reports may be graphical, some may be in a tabular format. A user may print the report through a standard Windows-style print dialog.

A user may view past performance of the Neugent (e.g., actual versus predicted). The statistical summaries may be stored by the Evaluation module of the Manager Service, and therefore may not need to be recalculated. A user also may print the report. Access restrictions to sensitive information may be provided through assignment of user account system security levels.

A Web Historian service may optionally be provided to provide viewing of reports through a web browser (e.g., Internet Explorer).

Similar to the configuration utility, the Historian may have the same look-and-feel as other AHD interfaces. Statistical summaries and predictions may be stored by the Evaluation module of the Manager Service, and therefore may not need to be recalculated.

The Historian may display and print different reports for CallLevelNeugents and SLA Violation Neugents.

For SLA Violation Neugents, a number of reports may be available. Exemplary SLA Violation reports are illustrated

in FIGS. 25-27. For example, a bar-chart graph of predictions for open issues (e.g., Calls, ChangeOrders, etc.) may be provided. Various views of the report may be available, which group the issues by assignee, queue, time-to-violation, and SLA type. Hotspots may be provided on the bar-chart, which when clicked brings up a listbox of the issues contained within a prediction bar. Clicking on an issue within the listbox brings up further details of the issue contained within AHD. For AHD installations equipped with an AHD WebServer, clicking on AHDView on the report brings up full details on the issue within the AHD web application.

Another report for SLAViolation Neugents may show the regular evaluations for the Neugent instance. The evaluations are performed by testing the active Neugent model on closed data, withheld from training.

In addition, another report for SLA violation Neugents may show the accuracy of past predictions, measured as calls which were predicted are closed over time. The performance measures may include Event Coverage (e.g., the percentage of actual SLA Violations which were correctly identified by AHD, and Excess Coverage (e.g., the percentage of non-violations which AHD tagged as likely violations).

For CallLevel Neugents, various reports also may be available. Some exemplary CallLevel reports are illustrated in FIGS. 28 and 29. For example, one report may show the individual predictions made by a CallLevel Neugent. Another report may be similar to the first, except multiple predictions may be superimposed on one graph. In addition, another report may show the actual versus predicted, called Individual Performance, levels for individual predictions. Another report may show average

accuracy of all past predictions for a selected CallLevel Neugent.

Additional utilities

5 The following functions may be available as function calls in one or more dynamic link libraries (DLLs). GetSecurityLevel is a utility which obtains the AHD Security Level of the logged-on user which may be used to restrict usage of the Historian and Configuration 10 utilities. RaiseAHDEvent is a utility which passes alerts to the normal AHD Event handling channel when conditions warrant, such as when an SLA violation is predicted. The function may accept parameters (if needed) to qualify an event to be raised. An event is raised when the Neugent 15 has determined that an SLA Violation has been predicted. Upon startup of the Neugent, the interface is initialized with AHD so that message flow is enabled.

Whenever the Neugent predicts that an SLA Violations is likely, a PredictSLA method is invoked to update the 20 specified Request or Change Order. PredictSLA takes a number of actions, including the following: (a) check the specified Request or Change Order; (b) increment the predict_sla_violated value in the request or Change Order record; (c) add an activity log with the specified contents 25 of the Neugent message string.

Parallel/Distributed Processing

Multithreading may be used to enable parallel processing, when possible and appropriate. In addition, 30 AHD may distribute Neugents between multiple consoles, in order to improve performance or to facilitate scaling. For example, the individual modules may be distributed among two or more processors, to improve system performance and

scalability. For example, in a 24x7 operation there might be no real off-hour period which may be dedicated to model retraining. A list of server names and/or IP addresses may be maintained to facilitate interprocess communications.

5 Alternatively, the AHD service may run entirely on a dedicated Management Console, to minimize its impact on other subsystems at a customer site.

10 The above specific embodiments are illustrative, and many variations can be introduced on these embodiments without departing from the spirit of the disclosure or from the scope of the appended claims. Elements and/or features of different illustrative embodiments may be combined with each other and/or substituted for each other within the scope of this disclosure and appended claims.

15 Additional variations may be apparent to one of ordinary skill in the art from reading the following applications, which are incorporated herein by reference:

(a) U.S. Serial No. 60/374,064, filed April 19, 2002 and entitled "PROCESSING MIXED NUMERIC AND/OR NON-NUMERIC 20 DATA";

(b) U.S. Serial No. 60/374,020, filed April 19, 2002 and entitled "AUTOMATIC NEURAL-NET MODEL GENERATION AND MAINTENANCE";

25 (c) U.S. Serial No. 60/374,024, filed April 19, 2002 and entitled "VIEWING MULTI-DIMENSIONAL DATA THROUGH HIERARCHICAL VISUALIZATION";

(d) U.S. Serial No. 60/374,041, filed April 19, 2002 and entitled "METHOD AND APPARATUS FOR DISCOVERING EVOLUTIONARY CHANGES WITHIN A SYSTEM";

30 (e) U.S. Serial No. 60/373,977, filed April 19, 2002 and entitled "AUTOMATIC MODEL MAINTENANCE THROUGH LOCAL NETS";

(f) U.S. Serial No. 60/373,780, filed April 19, 2002

Docket No. 66653-A
20000175

and entitled "USING NEURAL NETWORKS FOR DATA MINING"; and
(g) PCT International Application No. PCT/US02/22977,
filed July 18, 2002 and entitled "METHOD AND APPARATUS FOR
DISCOVERING EVOLUTIONARY CHANGES WITHIN A SYSTEM".

5